(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平10－198642

(43)公開日　平成10年(1998)７月31日

(51)Int.Cl.[6]		識別記号	FI		
Ｇ０６Ｆ	15/16	３７０	Ｇ０６Ｆ	15/16	３７０Ｎ
	13/00	３５７		13/00	３５７Ｚ

審査請求　未請求　請求項の数９　ＯＬ　（全 14 頁）

(21)出願番号　特願平9－1970

(22)出願日　平成９年(1997)１月９日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号

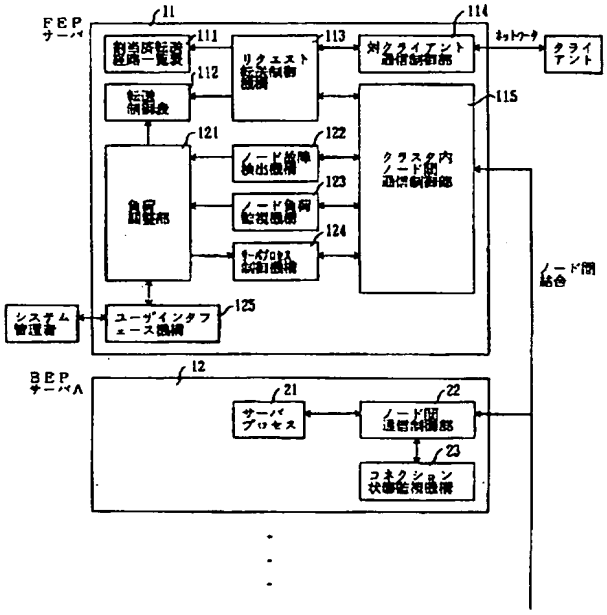
(72)発明者　大橋　勝之
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

(74)代理人　弁理士　山谷　晧榮　（外２名）

(54)【発明の名称】　サーバ装置

(57)【要約】
【課題】１種類のサービスの処理を複数台のバックエンドサーバで負荷分散して処理性能を向上すること。
【解決手段】クライアントのリクエストを処理するサーバプロセス２１を配置した複数個のバックエンドサーバ１２と、クライアントからのリクエストを受信するフロントエンドサーバ１１と、該フロントエンドサーバ１１に配置し、前記受信したリクエストを適切な前記バックエンドサーバ１２に転送するリクエスト転送制御機構１１３とを備え、該リクエスト転送制御機構１１３は、クライアントの識別情報を用いて、同一サービスに対する複数のクライアントからのリクエストを転送する前記バックエンドサーバ１２を決定してリクエストを転送する。

本発明の原理説明図



1

【特許請求の範囲】

【請求項１】クライアントのリクエストを処理するサーバプロセスを配置した複数個のバックエンドサーバと、クライアントからのリクエストを受信するフロントエンドサーバと、

該フロントエンドサーバに配置し、前記受信したリクエストを適切な前記バックエンドサーバに転送するリクエスト転送制御機構とを備え、

該リクエスト転送制御機構は、クライアントの識別情報を用いて、同一サービスに対する複数のクライアントからのリクエストを転送する前記バックエンドサーバを決定してリクエストを転送することを特徴としたサーバ装置。

【請求項２】前記リクエスト転送制御機構は、リクエスト転送制御表を用いて転送するリクエストの比率をバックエンドサーバ毎に制御することを特徴とした請求項１記載のサーバ装置。

【請求項３】サーバの故障を検出するノード故障検出機構を設け、

該ノード故障検出機構がサーバの故障を検出すると、リクエスト転送制御機構は、故障したバックエンドサーバへのリクエスト転送を中止して、クライアントからの再リクエスト時に正常運用している別のバックエンドサーバにリクエストを転送することを特徴とした請求項１又は２記載のサーバ装置。

【請求項４】前記バックエンドサーバの負荷を監視してリクエストの転送比率を変更するノード負荷監視機構を設け、

該ノード負荷監視機構が負荷が高い前記バックエンドサーバを発見すると、リクエスト転送制御機構は、負荷の高いバックエンドサーバへの転送比率を下げて、前記バックエンドサーバ間の負荷を平均化することを特徴とした請求項１～３のいずれかに記載のサーバ装置。

【請求項５】システム管理者が前記バックエンドサーバへの転送比率を制御するためのユーザインタフェース機構を設けることを特徴とした請求項１～４のいずれかに記載のサーバ装置。

【請求項６】前記サーバプロセスとクライアントプロセス間のコネクション状態を監視するコネクション状態監視機構を設け、

前記リクエスト転送制御機構は、前記コネクション状態監視機構にコネクション状態を問い合わせ、サービス中のコネクションに対しては、該サービス終了後に、転送経路の変更を行うことを特徴とした請求項１～５のいずれかに記載のサーバ装置。

【請求項７】前記サーバプロセスの制御を行うサーバプロセス制御機構を設け、

あるサービスのサーバプロセスが配置された各バックエンドサーバの負荷が高くなった時に、前記サーバプロセス制御機構はそのサーバプロセスが未配置なバックエンドサーバにサーバプロセスを起動することを特徴とした請求項１～６のいずれかに記載のサーバ装置。

2

【請求項８】前記リクエスト転送制御機構を、ネットワークの通信制御をするネットワークドライバとパケットの処理をするオペレーティングシステムのパケット処理部の中間に配置したパケットフィルタで構成することを特徴とした請求項１～７のいずれかに記載のサーバ装置。

【請求項９】前記クライアントの識別情報として、クライアントから受信したリクエストのパケットのソースアドレスとソースポート番号のペアを使用することを特徴とした請求項１～８のいずれかに記載のサーバ装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、計算機環境におけるクライアント／サーバシステムのサーバの構成に関し、特に、サーバをクラスタシステム（複数のノードを結合したシステム）で構成する際のノード（サーバ）間の負荷分散を行うサーバ装置に関する。

【０００２】近年、インターネット等の開放型システムの普及に伴い、クライアント／サーバシステムへの関心が高まり、特に不特定多数のクライアントからの要求が集中することが予想されるサーバシステムを高性能／高可用化する（利用可能率を高くする）ための技術として、クラスタシステムの負荷分散に関する技術の開発が望まれている。

【０００３】

【従来の技術】従来のクラスタシステムにおける負荷分散技術の１つに、ルータやパケットフィルタによるリダイレクション（仕事の割り振り）がある。この方式では、サーバはクライアントからのリクエストを受けるフロントエンドノード（ＦＥＰノード）と、クライアントからのリクエストを処理する機能を複数台のノードに分散配置したバックエンドノード（ＢＥＰノード）群から構成されていた。全てのクライアントはＦＥＰノードに対してリクエストを送り、ＦＥＰノードは要求されたサービスに対して適切なＢＥＰノードにリクエストを転送して負荷分散を行っていた。

【０００４】図７は従来例の説明図（１）、図８は従来例の説明図（２）である。以下図７、図８に基づいて従来例の説明をする。図７において、従来のリダイレクションを利用した負荷分散の例であり、サーバドメイン１と複数のクライアントドメイン①、②がＷＡＮ（ワイドエリアネットワーク）を介して接続されている。

【０００５】サーバドメイン１には、ＦＥＰサーバ１１、複数のサーバ１２（サーバＡ～Ｃ）、ドメインネイムサーバ（ＤＮＳ）１３、ルータ１４が設けてある。クライアントドメイン①には、複数のクライアント（１）～（３）、ＤＮＳ４３、ルータ４４が設けてある。クライアントドメイン②には、複数のクライアント（４）～

（6）、DNS43、ルータ44が設けてある。

【0006】FEPサーバ11は、LAN（ローカルエリアネットワーク）でクライアントからのリクエストを受信し、高速ノード間結合により適切なBEPノードにリクエストを転送する転送制御プロセスを有するFEPノードである。サーバ12（サーバA～C）は、それぞれ異なる種類のサービスX～Zを処理するサーバプロセスX～Zを有するBEPノードである。

【0007】DNS13は、LANと接続されておりクライアント等から問い合わせのあったサービス名に対応するIPアドレス（各コンピュータ固有のアドレス）を応答するものである。例えば、図8（b）はIPアドレス例の説明であり、図8（b）のようにクライアントは、DNS13からサービス名「ftp-server」のIPアドレス「133.160.12.5」を取得して、この取得したIPアドレスに向かってリクエストを出すものである。ルータ14は、WANとLAN間の中継処理を行うものである。

【0008】複数のクライアント（1）～（3）、複数のクライアント（4）～（6）は、それぞれのLANで接続されておりサービス（データ処理）を要求する側である。DNS43は、LANと接続されており問い合わせのあったサービス名に対応するIPアドレス応答するものである。ルータ44は、LANとWAN間の中継処理を行うものである。

【0009】図8（a）は仕事の割り振り表の説明であり、FEPノード上のプロセスによって管理される各サーバプロセスの配置を示すものである。図8（a）において、サービス識別子に対応する転送先サーバ識別子が示されている。例えば、サービスXはサーバAに、サービスYはサーバBに、サービスZはサーバCに転送することが示されている。

【0010】全てのクライアントからのリクエストは、クライアントドメインのLAN、ルータ44、WAN、ルータ14、サーバドメイン1のLANを介してFEPノードに送られ、FEPノードは、要求されたサービスの種類に応じてリクエストの転送先ノードを決定し、適切なBEPノードに、高速ノード間結合を介してリクエストを転送する。BEPノードA～C上のサーバプロセスX～Zは、FEPノードから転送されたリクエストを処理してクライアントにリプライするものであった。

【0011】

【発明が解決しようとする課題】前記従来のリダイレクション方式においては、BEPノード毎に異なるサービスを処理するサーバプロセスを配置し、図8（a）のような表を用いて、クライアントが要求したサービスをキーとして転送先BEPノードを決定していた。このため、1種類のサービスの処理が多くなっても複数台のBEPノードで負荷分散することができない課題があった。

【0012】本発明は、このような従来の課題を解決し、同一サービスに対する複数のクライアントからのリクエストを複数台のサーバノードで負荷分散をできるようにし、また、自動的にサーバノード故障隠蔽機能や、負荷調整機能を持つことにより、システム管理コストを削減することを目的とする。

【0013】

【課題を解決するための手段】図1は本発明の原理説明図である。図1中、11はフロントエンドサーバ、12はバックエンドサーバ、21はサーバプロセス、22はノード間通信制御部、23はコネクション状態監視機構、111は割当済転送経路一覧表、112はリクエスト転送制御表、113はリクエスト転送制御機構、114は対クライアント通信制御部、115はクラスタ内ノード間通信制御部、121は負荷調整部、122はノード故障検出機構、123はノード負荷監視機構、124はサーバプロセス制御機構、125はユーザインタフェース機構である。

【0014】本発明は前記従来の課題を解決するため次のように構成した。

（1）：クライアントのリクエストを処理するサーバプロセス21を配置した複数個のバックエンドサーバ12と、クライアントからのリクエストを受信するフロントエンドサーバ11と、該フロントエンドサーバ11に配置し、前記受信したリクエストを適切な前記バックエンドサーバ12に転送するリクエスト転送制御機構113とを備え、該リクエスト転送制御機構113は、クライアントの識別情報を用いて、同一サービスに対する複数のクライアントからのリクエストを転送する前記バックエンドサーバ12を決定してリクエストを転送する。

【0015】（2）：前記（1）のサーバ装置において、前記リクエスト転送制御機構113は、リクエスト転送制御表112を用いて転送するリクエストの比率をバックエンドサーバ12毎に制御する。

【0016】（3）：前記（1）又は（2）のサーバ装置において、サーバの故障を検出するノード故障検出機構122を設け、ノード故障検出機構122がサーバの故障を検出すると、リクエスト転送制御機構113は、故障したバックエンドサーバ12へのリクエスト転送を中止して、クライアントからの再リクエスト時に正常運用している別のバックエンドサーバ12にリクエストを転送する。

【0017】（4）：前記（1）～（3）のサーバ装置において、前記バックエンドサーバ12の負荷を監視してリクエストの転送比率を変更するノード負荷監視機構123を設け、該ノード負荷監視機構123が負荷が高い前記バックエンドサーバ12を発見すると、リクエスト転送制御機構113は、負荷の高いバックエンドサーバ12への転送比率を下げて、前記バックエンドサーバ12間の負荷を平均化する。

【００１８】（５）：前記（１）～（４）のサーバ装置において、システム管理者が前記バックエンドサーバ１２への転送比率を制御するためのユーザインタフェース機構１２５を設ける。

【００１９】（６）：前記（１）～（５）のサーバ装置において、前記サーバプロセス２１とクライアントプロセス間のコネクション状態を監視するコネクション状態監視機構２３を設け、前記リクエスト転送制御機構１１３は、前記コネクション状態監視機構２３にコネクション状態を問い合わせ、サービス中のコネクションに対しては、該サービス終了後に、転送経路の変更を行う。

【００２０】（７）：前記（１）～（６）のサーバ装置において、前記サーバプロセス２１の制御を行うサーバプロセス制御機構１２４を設け、あるサービスのサーバプロセス２１が配置された各バックエンドサーバ１２の負荷が高くなった時に、前記サーバプロセス制御機構１２４がそのサーバプロセス２１が未配置なバックエンドサーバ１２にサーバプロセス２１を起動する。

【００２１】（８）：前記（１）～（７）のサーバ装置において、前記リクエスト転送制御機構１１３を、ネットワークの通信制御をするネットワークドライバとパケットの処理をするオペレーティングシステムのパケット処理部の中間に配置したパケットフィルタで構成する。

【００２２】（９）：前記（１）～（８）のサーバ装置において、前記クライアントの識別情報として、クライアントから受信したリクエストのパケットのソースアドレスとソースポート番号のペアを使用する。

【００２３】（作用）前記構成に基づく作用を説明する。複数個のバックエンドサーバ１２に配置したサーバプロセス２１でクライアントのリクエストを処理し、フロントエンドサーバ１１に配置したリクエスト転送制御機構１１３で受信したクライアントからのリクエストをクライアントの識別情報を用いて、同一サービスに対する複数のクライアントからのリクエストを転送する前記バックエンドサーバ１２を決定してリクエストを転送する。このため、１種類のサービスの処理を複数台のバックエンドサーバで負荷分散でき、処理性能を向上することができる。

【００２４】また、前記リクエスト転送制御機構１１３で、リクエスト転送制御表１１２を用いて転送するリクエストの比率をバックエンドサーバ１２毎に制御する。このため、バックエンドサーバの処理能力に応じた負荷分散ができ、より処理性能を向上することができる。

【００２５】さらに、ノード故障検出機構１２２がサーバの故障を検出すると、リクエスト転送制御機構１１３で、故障したバックエンドサーバ１２へのリクエスト転送を中止して、クライアントからの再リクエスト時に正常運用している別のバックエンドサーバ１２にリクエストを転送する。このため、クライアントからの再リクエストにより、バックエンドサーバの故障をクライアントから隠蔽することができる。

【００２６】また、ノード負荷監視機構１２３が負荷が高い前記バックエンドサーバ１２を発見すると、リクエスト転送制御機構１１３で、負荷の高いバックエンドサーバ１２への転送比率を下げて、前記バックエンドサーバ１２間の負荷を平均化する。このため、常に負荷を平均化することができ、処理性能を向上することができる。

【００２７】さらに、ユーザインタフェース機構１２５で、システム管理者が前記バックエンドサーバ１２への前記転送比率を制御する。このため、例えば、活性保守のためのバックエンドサーバの転送比率をゼロとしてノードの一時的な切り離しや再組み込みが可能となる。

【００２８】また、前記リクエスト転送制御機構１１３は、コネクション状態監視機構２３にコネクション状態を問い合わせ、サービス中のコネクションに対しては、該サービス終了後に、転送経路の変更を行う。このため、サービス途中に新たなバックエンドサーバに切り換わることがなく、処理の中断を防止することができる。

【００２９】さらに、あるサービスのサーバプロセス２１が配置された各バックエンドサーバ１２の負荷が高くなった時に、サーバプロセス制御機構１２４がそのサーバプロセス２１が未配置なバックエンドサーバ１２にサーバプロセス２１を起動する。このため、複数個のバックエンドサーバ全体の負荷を調整することができ、処理性能を向上することができる。

【００３０】また、前記リクエスト転送制御機構１１３を、ネットワークの通信制御をするネットワークドライバとパケットの処理をするオペレーティングシステムのパケット処理部の中間に配置したパケットフィルタで構成する。このため、パケットフィルタで１種類のサービスの処理を複数台のバックエンドサーバで負荷分散でき、処理性能が向上する。

【００３１】さらに、前記クライアントの識別情報として、クライアントから受信したリクエストのパケットのソースアドレスとソースポート番号のペアを使用する。このため、クライアントプロセス毎にリクエストの転送先ノードの制御を行うことができる。

【００３２】

【発明の実施の形態】図２～図６は本発明の実施の形態を示した図である。本発明では、ＦＥＰノードにおけるリクエスト転送制御に、クライアントが要求したサービスの種類に加えて、リクエストを送ったクライアント識別子を用い、同一サービスに対するリクエストの転送先ＢＥＰノードをクライアント毎に制御するものである。以下、図面に基づいて本発明の実施の形態を説明する。

【００３３】（１）：サーバ装置の説明

①：装置構成の説明

図２は装置構成図（１）である。以下、図２に基づいてサーバ装置の説明をする。

【００３４】図２において、サーバ装置は、ＦＥＰサーバ（ＦＥＰノード）１１と複数のＢＥＰサーバ（ＢＥＰノード）１２ａ、１２ｂ、１２ｃ、・・・がノード間結合３で接続されている。クライアント４は、ＬＡＮあるいはＷＡＮ５を経由してＦＥＰノード１１と結合されている。

【００３５】ＦＥＰサーバ１１には、割当済転送経路一覧表１１１、転送制御表１１２、リクエスト転送処理部１１３、対クライアント通信制御部１１４、クラスタ内ノード間通信制御部１１５、負荷調整部１２１、ノード故障検出部１２２、ノード負荷監視部１２３、サーバプロセス制御部１２４、システム管理者インタフェース１２５が設けてある。

【００３６】ＢＥＰサーバ１２ａには、サーバプロセス２１ａ、ノード間通信制御部２２ａ、コネクション状態監視部２３ａが設けてあり、ＢＥＰサーバ１２ｂには、サーバプロセス２１ｂ、ノード間通信制御部２２ｂ、コネクション状態監視部２３ｂが設けてあり、ＢＥＰサーバ１２ｃにも、サーバプロセス２１ｃ、ノード間通信制御部２２ｃ、コネクション状態監視部２３ｃが設けてある。

【００３７】割当済転送経路一覧表１１１は、クライアント毎に割り当てたリクエストの転送経路を記録したものである。転送制御表１１２は、リクエストを処理するサーバプロセスの配置とリクエストの振り分け比率を記録したものである。リクエスト転送処理部１１３は、クライアントからのリクエストをＢＥＰノードに転送するものである。対クライアント通信制御部１１４は、クライアントとの通信を制御するものである。クラスタ内ノード間通信制御部１１５は、ＢＥＰノードとの通信を制御するものである。負荷調整部１２１は、システム管理者の要求や発生したイベントに応じて動的に転送制御を調整するものである。ノード故障検出部１２２は、ＢＥＰノードの故障を検出するものである。ノード負荷監視部１２３は、ＢＥＰノードの負荷状態を監視するものである。サーバプロセス制御部１２４は、ＢＥＰノードに必要なサーバプロセスを起動するものである。システム管理者インタフェース１２５は、システム管理者６にユーザインタフェースを提供するものである。

【００３８】サーバプロセス２１ａ、２１ｂ、２１ｃは、クライアントからのリクエストを処理するプロセスである。ノード間通信制御部２２ａ、２２ｂ、２２ｃは、ＦＥＰノードとの通信を制御するものである。コネクション状態監視部２３ａ、２３ｂ、２３ｃは、クライアントとサーバプロセスとのコネクション状態を監視するものである。

【００３９】②：転送制御表と割当済転送経路一覧表の説明

図３は転送制御表と割当済転送経路一覧表の説明図であり、図３（ａ）は転送制御表の説明、図３（ｂ）は割当済転送経路一覧表の説明である。

【００４０】図３（ａ）において、転送制御表１１２には、リクエストの宛て先としてのポート番号等を使用するサービス識別子、転送先となる利用可能なＢＥＰノードのＩＰアドレス等を使用する利用可能ＢＥＰノード識別子、同一サービスに対するリクエストの分配比率（転送比率）である比率が設けてある。

【００４１】図３（ｂ）において、割当済転送経路一覧表１１１には、リクエストの宛て先としてのポート番号等を使用するサービス識別子、ソースＩＰアドレスとソースポートのポート番号等を使用するクライアント識別子、転送先のＢＥＰノードのＩＰアドレス等を使用する転送先ＢＥＰノード識別子が設けてある。

【００４２】（２）：サーバ装置の動作の説明

以下、図２に基づいてサーバ装置の動作を説明する。

①：クライアントからリクエストがあった場合の説明

クライアント４がリクエストをＬＡＮあるいはＷＡＮ５を経由してＦＥＰノード１１に送ると、対クライアント通信制御部１１４によって受信されたリクエストはリクエスト転送処理部１１３に送られる。リクエスト転送処理部１１３は、まず、割当済転送経路一覧表１１１を用いて、すでに転送経路を割り当てたことがあるクライアントからのリクエストかどうか調べる。この割当済転送経路一覧表１１１には、リクエストしたクライアントの識別子と要求したサービスの識別子のペアーをキーとして、割り当てられた転送先ＢＥＰノードの識別子が記録されている。

【００４３】割当済転送経路一覧表１１１にクライアントのエントリがない場合には、リクエスト転送処理部１１３は、転送制御表１１２を参照して、転送先ＢＥＰノードを決定する。この転送制御表１１２には、利用可能ＢＥＰノード識別子とリクエストの分配比率がサービス毎に記録されている。

【００４４】転送制御表１１２でリクエストの転送先ＢＥＰノードが決定すると、リクエスト転送処理部１１３は、割当済転送経路一覧表１１１に転送先を割り当てたクライアントのエントリを作り、クラスタ内ノード間通信制御部１１５、ノード間結合３を経由して、リクエストを転送先として決定したＢＥＰノードに転送する。

【００４５】このリクエストが転送されたＢＥＰノード（例えば、ＢＥＰノード１２ａとする）では、ノード間通信制御部２２ａがリクエストを受信してサーバプロセス２１ａにリクエストを送る。サーバプロセス２１ａは、リクエストを処理すると、ノード間通信制御部２２ａ、ノード間結合３を経由して、ＦＥＰノード１１にリプライを送る。ＦＥＰノード１１は、クラスタ内ノード間通信制御部１１５でリプライを受け、リクエスト転送処理部１１３が対クライアント通信制御部１１４、ＬＡＮあるいはＷＡＮ５を介してクライアント４にリプライを返す。

【００４６】割当済転送経路一覧表１１１にエントリが存在するクライアントからのリクエストの場合は、リクエスト転送処理部１１３は、割当済転送経路一覧表１１１で割り当てられているＢＥＰノードにリクエストを送ることになる。このような処理を行うことによって、同一のクライアント識別子で識別されるクライアントからのリクエストは同一のＢＥＰサーバに転送されることになる。

【００４７】②：ノード故障の場合の説明

ノード故障検出部１２２は、定期的に全てのＢＥＰノードが正常動作していることを確認する。ノード故障検出部１２２は、ＢＥＰノードの故障を検出すると、負荷調整部１２１を経由して転送制御表１１２を書き換えて、故障したＢＥＰノードへのリクエストの分配を停止する。リクエスト転送処理部１１３は、割当済転送経路一覧表１１１内の転送先ＢＥＰノードが故障したＢＥＰノードとなっている全てのエントリを削除する。

【００４８】このＢＥＰノードの故障によって、故障したＢＥＰノードを利用していたクライアントとＢＥＰノード上のサーバプロセスとのコネクションは切断されるが、クライアントがリクエストをリトライすることによって故障は隠蔽される。

【００４９】クライアントからの再リクエストはリクエスト転送処理部１１３に送られると、リクエスト転送処理部１１３は、割当済転送経路一覧表１１１を調べるが、再リクエストしたクライアントのエントリは既に削除されているため、リクエスト転送処理部１１３は、転送制御表１１２に基づいて新たな転送先を決定してリクエストを転送する。リクエストを転送されたＢＥＰノードでは、通常のリクエストと同様にリクエストを処理してリプライする。

【００５０】③：ノード負荷監視の説明

ノード負荷監視部１２３は、定期的に全てのＢＥＰノードの負荷状態を監視する。ノード負荷監視部１２３は、負荷が高いＢＥＰノードを発見すると、負荷調整部１２１を経由して転送制御表１１２を書き換えて、負荷が高いＢＥＰノードへのリクエストの分配比率を下げる。リクエスト転送処理部１１３は、分配比率変更時に割り当て済の転送経路の状態を各ＢＥＰサーバのコネクション状態監視部２３ａ、２３ｂ、２３ｃ、・・・に問い合わせ、割当済転送経路一覧表１１１からアクティブでないコネクションの転送経路のエントリを削除し、その時点のリクエスト分配比率を算出する。

【００５１】リクエスト転送処理部１１３は、以降の新規クライアントからのリクエスト転送先の調整や、コネクション状態監視部２３ａ、２３ｂ、２３ｃ、・・・に問い合わせによる割当済転送経路一覧表１１１の更新により、書き換えられた転送制御表１１２で指定された分配比率に実際の分配比率を近付けるようにする。

【００５２】④：サーバプロセスの起動・停止の説明

サーバプロセス制御部１２４は、クラスタ内ノード間通信制御部１１５、ノード間結合３を介して各ＢＥＰノード上のサーバプロセスを制御するものである。

【００５３】あるサービスに割り当てられた全てのＢＥＰノードの負荷が高い場合には、ノード負荷監視部１２３は負荷調整部１２１にＢＥＰノードの割り当て増を依頼する。各サービスへのＢＥＰノードの割り当てを管理している負荷調整部１２１は、負荷が高いサービスへのＢＥＰノード割り当ての増を決めると、新たに割り当てたＢＥＰノードへサーバプロセスの起動をサーバプロセス制御部１２４に依頼する。そこで、サーバプロセス制御部１２４は、クラスタ内ノード間通信制御部１１５、ノード間結合３を経由して、指定されたＢＥＰノード上のサーバプロセスを起動する。このサーバプロセスの起動が完了すると、負荷調整部１２１は、新たに割り当てたＢＥＰノードにリクエストを転送するように転送制御表１１２を書き換える。

【００５４】新たに割り当てたＢＥＰノード上で、それまで動作していたサーバプロセスを停止させることが必要な場合、負荷調整部１２１は、転送制御表１１２を書き換えてリクエストに対する分配を停止する（分配比率を０とする）。リクエスト転送処理部１１３は、リクエストの転送の停止が指示されたＢＥＰノードのコネクション状態監視部にコネクションの状態を問い合わせ、割当済転送経路一覧表１１１の未使用状態のエントリを削除する。負荷調整部１２１は、割当済転送経路一覧表１１１を参照して、指定したＢＥＰノードへのコネクションのエントリが無くなった時点で、サーバプロセス制御部１２４にサーバプロセスの停止を依頼する。サーバプロセス制御部１２４は、クラスタ内ノード間通信制御部１１５、ノード間結合３を経由して、指定されたＢＥＰノード上のサーバプロセスを停止する。

【００５５】⑤：システム管理者インタフェースの説明

システム管理者インタフェース１２５は、システム管理者６にシステム管理のためのインタフェースを提供するものである。

【００５６】システム管理者インタフェース１２５が提供する機能としては、クラスタへの動的な（運用中の）ノードの追加／切り離し、サービスへのＢＥＰノード割り当て指定／変更、その他システム状態の参照等の機能がある。

【００５７】（３）：ＵＮＩＸ（ＡＴ＆Ｔ社の汎用のオペレーティングシステム）マシンによる説明

以下、図４～図６に従ってＵＮＩＸマシンによる説明をする。

【００５８】①：クライアント／サーバシステムの説明

図４はクライアント／サーバシステムの説明図である。以下図４に基づいて説明をする。

【００５９】図４において、リダイレクションを利用した負荷分散の例であり、サーバドメイン１と複数のクラ

イアントドメイン①、②がＷＡＮを介して接続されている。サーバドメイン１には、ＦＥＰサーバ１１、複数のサーバ１２（サーバＡ～Ｆ）、ドメインネイムサーバ（ＤＮＳ）１３、ルータ１４が設けてある。クライアントドメイン①には、複数のクライアント（１）～（５）、ＤＮＳ４３、ルータ４４が設けてある。クライアントドメイン②には、複数のクライアント（６）～（１０）、ＤＮＳ４３、ルータ４４が設けてある。

【００６０】ＦＥＰサーバ１１は、ＬＡＮ（ローカルエリアネットワーク）でクライアントからのリクエストを受信し、高速ノード間結合により適切なＢＥＰノード１２にリクエストを転送するパケットフィルタ１１３を有するＦＥＰノードである。サーバ１２（サーバＡ～Ｆ）は、サービスＸ、Ｙ、Ｚを処理するサーバプロセス２１（サーバプロセスＸ、Ｙ、Ｚ）を有するＢＥＰノードである。ＤＮＳ１３は、ＬＡＮと接続されておりクライアント等から問い合わせのあったサービス名に対応するＩＰアドレス（各コンピュータ固有のアドレス）を応答するものである。

【００６１】複数のクライアント（１）～（５）、複数のクライアント（６）～（１０）は、それぞれのＬＡＮで接続されておりサービス（データ処理）を要求する側である。ＤＮＳ４３は、ＬＡＮと接続されており問い合わせのあったサービス名に対応するＩＰアドレス応答するものである。ルータ４４は、ＬＡＮとＷＡＮ間の中継処理を行うものである。

【００６２】②：サーバ装置の説明

ＵＮＩＸマシンのサーバ装置（クラスタシステム）における負荷分散について、図２のリクエスト転送処理部１１３は、ネットワークのデータ処理をするストリーム処理部の最下層に位置するパケットフィルタとして実現される。そして、割当済転送経路一覧表１１１と転送制御表１１２はパケットフィルタ内に置かれている。また、図２の負荷調整部１２１はユーザ空間内のＵＮＩＸプロセスである負荷調整プロセスとして実現する。この負荷調整プロセスは、ＵＮＩＸ　ＯＳ（オペレーティングシステム）が提供するＩＯＣＴＬシステムコールを利用して、パケットフィルタ内の転送制御表の書き換えを行うものである。

【００６３】図５は装置構成図（２）である。以下、図５に基づいてサーバ装置の説明をする。図５において、サーバ装置は、ＦＥＰノード１１とＢＥＰノード１２（実際は複数のＢＥＰノードが設けられる）がノード間結合ネットワーク３で接続されている。クライアントとは、ＬＡＮ等のネットワーク５を介してＦＥＰノード１１と結合されている。

【００６４】ＦＥＰノード１１には、カーネル（中枢部分）空間とユーザ空間があり、カーネル空間にはストリーム処理部、外部ネットワークドライバ１１４、内部ネットワークドライバ１１５が設けてある。ストリーム処理部には、パケットフィルタ１１３、ＩＰパケット処理層１３１、ＴＣＰパケット処理層１３２が設けてあり、パケットフィルタ１１３には、割当済転送経路一覧表１１１と転送制御表１１２が設けてある。ユーザ空間には、負荷調整プロセス１２１、ノード故障検出プロセス１２２、ノード負荷監視プロセス１２３、サーバ制御プロセス１２４、ユーザＩ／Ｆ処理プロセス１２５が設けてある。

【００６５】ＢＥＰノード１２にも、カーネル空間とユーザ空間があり、カーネル空間にはストリーム処理部、内部ネットワークドライバ２２が設けてある。ストリーム処理部には、ＩＰパケット処理層２４１、ＴＣＰパケット処理層２４２が設けてある。ユーザ空間には、接続監視プロセス２３、サーバプロセス２１が設けてある。

【００６６】割当済転送経路一覧表１１１は、クライアント毎に割り当てたリクエストの転送経路を記録したものである。転送制御表１１２は、リクエストを処理するサーバプロセスの配置とリクエストの振り分け比率を記録したものである。パケットフィルタ１１３は、クライアントからのリクエストをＢＥＰノードに転送するものである。外部ネットワークドライバ１１４は、クライアントとの通信を制御するものである。内部ネットワークドライバ１１５は、ＢＥＰノードとの通信を制御するものである。負荷調整プロセス１２１は、システム管理者の要求や発生したイベントに応じて動的に転送制御を調整するものである。ノード故障検出プロセス１２２は、ＢＥＰノードの故障を検出するものである。ノード負荷監視プロセス１２３は、ＢＥＰノードの負荷状態を監視するものである。サーバ制御プロセス１２４は、ＢＥＰノードに必要なサーバプロセスを起動するものである。ユーザＩ／Ｆ処理プロセス１２５は、システム管理者にユーザインタフェースを提供するものである。ＩＰパケット処理層１３１は、通信プロトコルの一つであるインターネットプロトコル（ＩＰ）のパケットの処理を行うものである。ＴＣＰパケット処理層１３２は、通信プロトコルの一つであるＴＣＰ（transmission control protocol）のパケットの処理を行うものである。

【００６７】サーバプロセス２１は、クライアントからのリクエストを処理するプロセスである。内部ネットワークドライバ２２は、ＦＥＰノード１１との通信を制御するものである。接続監視プロセス２３は、クライアントとサーバプロセスとのコネクション状態を監視するものである。ＩＰパケット処理層２４１は、インターネットプロトコルのパケットの処理を行うものである。ＴＣＰパケット処理層２４２は、ＴＣＰのパケットの処理を行うものである。

【００６８】③：転送制御表と割当済転送経路一覧表の説明

図６は転送制御表と割当済転送経路一覧表の説明図であり、図６（ａ）は転送制御表の説明、図６（ｂ）は割当

済転送経路一覧表の説明である。

【0069】図6（a）において、転送制御表112には、サービス識別子、転送先BEPノード識別子、同一サービスに対するリクエストの分配比率である比率が設けてある。

【0070】サービス識別子としてデスティネーション（宛て先）ポートのポート番号を使用し、この例では「8080、8081、8082」が記録されている。転送先BEPノード識別子として、デスティネーションIPアドレスを使用し、この例では例えば、サービス「8080」のデスティネーションIPアドレスには「IP－A、IP－B」が記録され、これらの比率として、IP－Aには「1」、IP－Bには「1」と同じ比率が記録されている。

【0071】図6（b）において、割当済転送経路一覧表111には、サービス識別子、クライアント識別子、転送先BEPノード識別子が設けてある。サービス識別子としてデスティネーションポートのポート番号を使用し、この例では「8080、8081、8082」が記録されている。

【0072】クライアント識別子としてソースIPアドレスとソースポート番号の組を使用する。ソースIPアドレスはクライアントのIPアドレスであり、ソースポート番号はプロセス毎に設けられている。この例では、（例えば、図4のクライアント（1））ソースIPアドレス「IP－CL1」には3つのソースポート番号「PORT－CL1－xx1」、「PORT－CL1－xx2」、「PORT－CL1－xx3」が示されている。

【0073】転送先BEPノード識別子として、デスティネーションIPアドレスを使用し、この例では例えば、サービス「8080」のデスティネーションIPアドレスには「IP－A」と「IP－B」が記録されている。

【0074】なお、サービスとポート番号の対応はUNIX　OS上のファイルに定義され、クラスタ内の各ノードは、共通のサービスポートを使うものである。

④：サーバ装置の動作の説明

a：クライアントからリクエストがあった場合の説明

FEPノード11の外部ネットワークドライバ114は、クライアントからのリクエストを受信して、パケットフィルタ113に送る。パケットフィルタ113は、割当済転送経路一覧表111と転送制御表112を用いてパケットを転送するBEPノードを決定し、ノード間結合ネットワーク3を介してパケットをBEPノード12に送る。

【0075】このパケットが転送されたBEPノード12の内部ネットワークドライバ22は、受信したパケットをストリーム処理部に送る。サーバプロセス21には、UNIXが提供するシステムコールインタフェースを用いてTCPパケット処理層242に届いたパケットを読み出して、リクエストを処理する。

【0076】サーバプロセス21は、リクエストの処理が完了するとクライアントにリプライするために、UNIXのシステムコールを用いてパケットをストリーム処理部に渡す。このリプライのパケットは、内部ネットワークドライバ22、ノード間結合ネットワーク3を介してFEPノード11に転送された後、FEPノード11によって、クライアントに送られる。

【0077】b：ノード故障の場合の説明

ノード故障検出プロセス122は、ユーザプロセスとして実現されている。ノード故障検出プロセス122は、定期的に全てのBEPノードが正常動作していることを確認し、BEPノードの故障を検出すると、負荷調整プロセス121を経由してパケットフィルタ113の転送制御表112を書き換えて、故障したBEPノードへのリクエストの分配を停止する。パケットフィルタ113は、割当済転送経路一覧表111内の転送先BEPノードが故障したBEPノードとなっている全てのエントリを削除する。

【0078】このBEPノードの故障によって、故障したBEPノードを利用していたクライアントとBEPノード上のサーバプロセスとのコネクションは切断されるが、クライアントがリクエストをリトライすることによって故障は隠蔽される。

【0079】クライアントからの再リクエストはパケットフィルタ113に送られると、パケットフィルタ113は、割当済転送経路一覧表111を調べるが、再リクエストしたクライアントのエントリは既に削除されているため、パケットフィルタ113は、転送制御表112に基づいて新たな転送先を決定してリクエストを転送する。リクエストを転送されたBEPノードでは、通常のリクエストと同様にリクエストを処理してリプライする。

【0080】c：ノード負荷監視の説明

ノード負荷監視プロセス123は、定期的にBEPノード12と通信してBEPノード12の負荷状態を監視する。ノード負荷監視プロセス123は、負荷が高いBEPノード12を発見すると、負荷調整プロセス121に通知し、通知を受けた負荷調整プロセス121は、IOCTLシステムコールを用いてパケットフィルタ113内の転送制御表112を書き換える。

【0081】パケットフィルタ113は、BEPノード12上の接続監視プロセス23にクライアントの接続の状態を問い合わせる。接続監視プロセス23は、IOCTLを使って自ノードのストリーム処理部にコネクション状態を問い合わせ、パケットフィルタ113に通知する。パケットフィルタ113は、未使用のコネクションの転送先だけを変更して転送比率を調整する。

【0082】負荷調整プロセス121は、必要があれば、サーバ制御プロセス124にBEPノード12への

サーバプロセス２１の起動あるいは停止を依頼する。このように、本発明によれば、同一サービスに対する複数のクライアントからのリクエストを複数台のサーバノードで負荷分散することが可能となる。また、自動的なサーバノード故障隠蔽機能や、負荷調整機能を持つことにより、システム管理コストを削減することができる。

【００８３】

【発明の効果】以上説明したように、本発明によれば次のような効果がある。

（１）：フロントエンドサーバに配置したリクエスト転送制御機構で受信したクライアントからのリクエストをクライアントの識別情報を用いて、同一サービスに対する複数のクライアントからのリクエストを転送するバックエンドサーバを決定してリクエストを転送するため、１種類のサービスの処理を複数台のバックエンドサーバで負荷分散でき、処理性能を向上することができる。

【００８４】（２）：リクエスト転送制御機構で、リクエスト転送制御表を用いて転送するリクエストの比率をバックエンドサーバ毎に制御するため、バックエンドサーバの処理能力に応じた負荷分散ができ、より処理性能を向上することができる。

【００８５】（３）：ノード故障検出機構がサーバの故障を検出すると、リクエスト転送制御機構で、故障したバックエンドサーバへのリクエスト転送を中止して、クライアントからの再リクエスト時に正常運用している別のバックエンドサーバにリクエストを転送するため、クライアントからの再リクエストにより、バックエンドサーバの故障をクライアントから隠蔽することができる。

【００８６】（４）：ノード負荷監視機構が負荷が高いバックエンドサーバを発見すると、リクエスト転送制御機構で、負荷の高いバックエンドサーバへの転送比率を下げて、前記バックエンドサーバ間の負荷を平均化するため、常に負荷を平均化することができ、処理性能を向上することができる。

【００８７】（５）：ユーザインタフェース機構で、システム管理者がバックエンドサーバへの転送比率を制御するため、例えば、活性保守のためのバックエンドサーバの転送比率をゼロとしてノードの一時的な切り離しや再組み込みが可能となる。

【００８８】（６）：リクエスト転送制御機構は、コネクション状態監視機構にコネクション状態を問い合わせ、サービス中のコネクションに対しては、該サービス終了後に、転送経路の変更を行うため、サービス途中に新たなバックエンドサーバに切り換わることがなく、処理の中断を防止することができる。

【００８９】（７）：あるサービスのサーバプロセスが配置された各バックエンドサーバの負荷が高くなった時に、サーバプロセス制御機構がそのサーバプロセスが未配置なバックエンドサーバにサーバプロセスを起動するため、複数個のバックエンドサーバ全体の負荷を調整することができ、処理性能を向上することができる。

【００９０】（８）：リクエスト転送制御機構を、ネットワークの通信制御をするネットワークドライバとパケットの処理をするオペレーティングシステムのパケット処理部の中間に配置したパケットフィルタで構成するため、パケットフィルタで１種類のサービスの処理を複数台のバックエンドサーバで負荷分散でき、処理性能が向上する。

【００９１】（９）：クライアントの識別情報として、クライアントから受信したリクエストのパケットのソースアドレスとソースポート番号のペアを使用するため、クライアントプロセス毎にリクエストの転送先ノードの制御を行うことができる。

【図面の簡単な説明】

【図１】本発明の原理説明図である。

【図２】実施の形態における装置構成図（１）である。

【図３】実施の形態における転送制御表と割当済転送経路一覧表の説明図である。

【図４】実施の形態におけるクライアント／サーバシステムの説明図である。

【図５】実施の形態における装置構成図（２）である。

【図６】実施の形態における転送制御表と割当済転送経路一覧表の説明図である。

【図７】従来例の説明図（１）である。

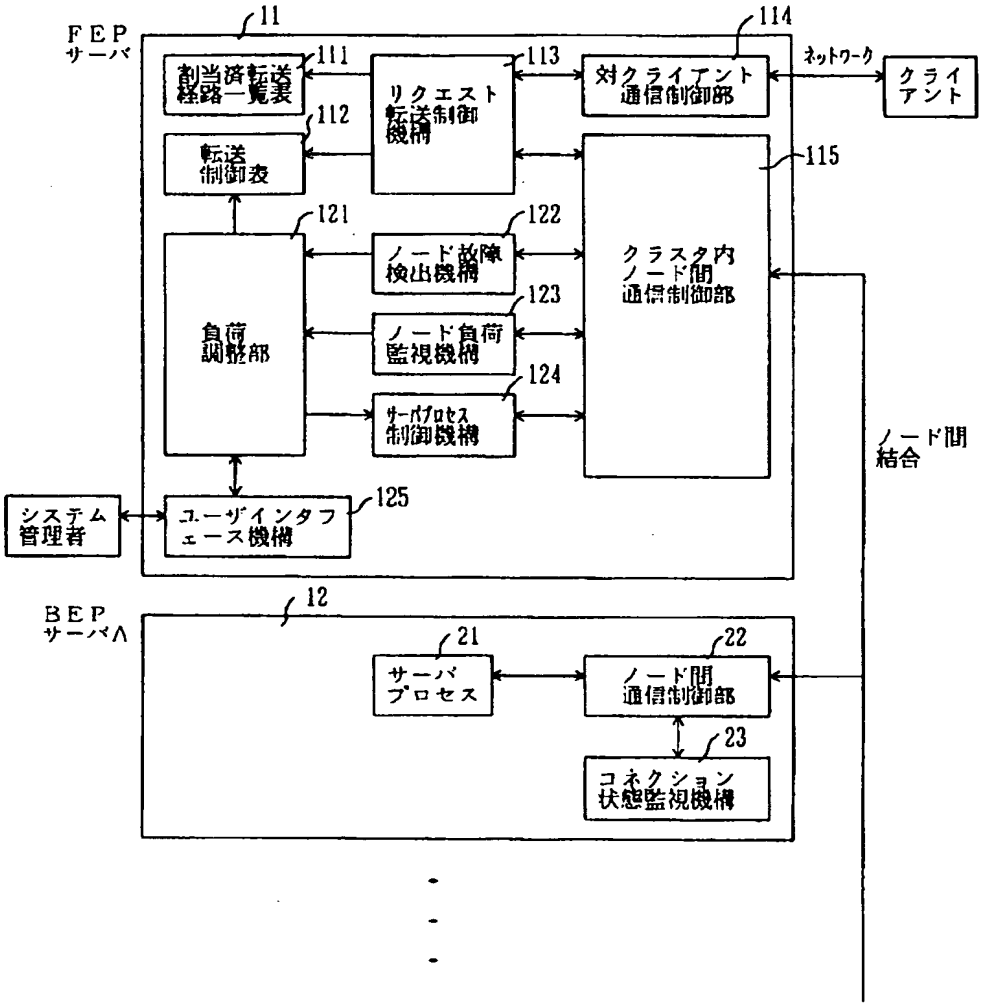
【図８】従来例の説明図（２）である。

【符号の説明】

１１　フロントエンドサーバ
１２　バックエンドサーバ
２１　サーバプロセス
２２　ノード間通信制御部
２３　コネクション状態監視機構
１１１　割当済転送経路一覧表
１１２　リクエスト転送制御表
１１３　リクエスト転送制御機構
１１４　対クライアント通信制御部
１１５　クラスタ内ノード間通信制御部
１２１　負荷調整部
１２２　ノード故障検出機構
１２３　ノード負荷監視機構
１２４　サーバプロセス制御機構
１２５　ユーザインタフェース機構

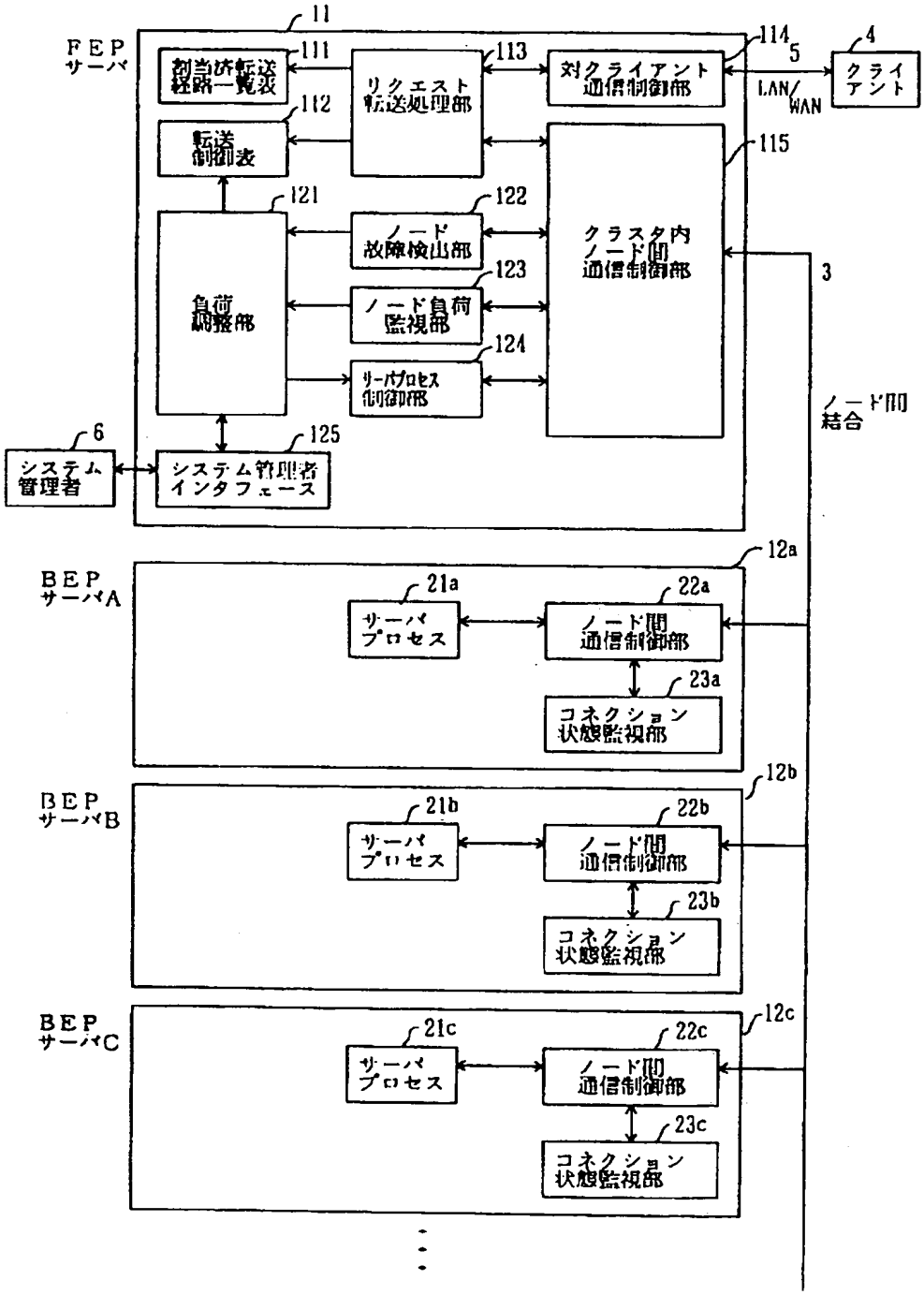
【図１】

本発明の原理説明図



【図２】

装置構成図(1)


FEP
サーバ
11
初当該転送
経路一覧表
111
転送
制御表
112
リクエスト
転送処理部
113
対クライアント
通信制御部
114
5
LAN/
WAN
4
クライ
アント
115
121
負荷
調整部
ノード
故障検出部
122
ノード負荷
監視部
123
リーバプロセス
制御部
124
クラスタ内
ノード間
通信制御部
3
ノード間
結合
6
システム
管理者
125
システム管理者
インタフェース
BEP
サーバA
12a
21a
サーバ
プロセス
22a
ノード間
通信制御部
23a
コネクション
状態監視部
BEP
サーバB
12b
21b
サーバ
プロセス
22b
ノード間
通信制御部
23b
コネクション
状態監視部
BEP
サーバC
12c
21c
サーバ
プロセス
22c
ノード間
通信制御部
23c
コネクション
状態監視部

【図３】

転送制御表と割当済転送経路一覧表の説明図

（ａ）転送制御表の説明

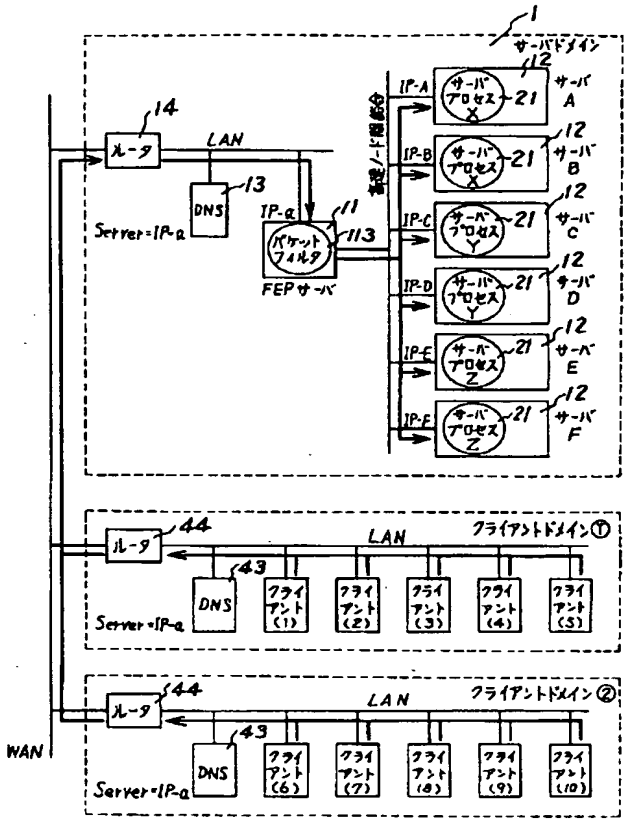
サービス識別子	利用可能ＢＥＰノード識別子	比　率

（ｂ）割当済転送経路一覧表の説明

サービス識別子	クライアント識別子	転送先ＢＥＰノード識別子

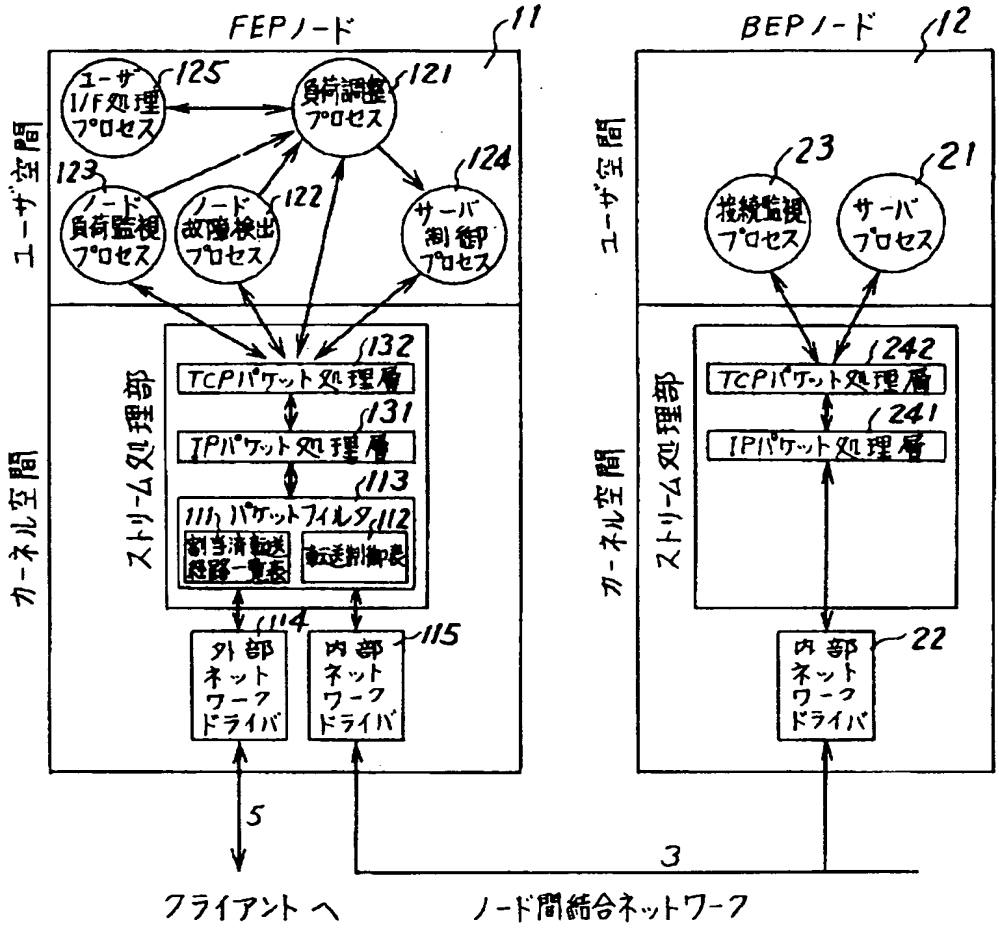
【図４】

クライアント/サーバシステムの説明図



【図５】

装置構成図（２）


FEPノード
11
BEPノード
12
ユーザ空間
カーネル空間
ユーザI/F処理プロセス
125
負荷調整プロセス
121
ノード負荷監視プロセス
123
ノード故障検出プロセス
122
サーバ制御プロセス
124
接続監視プロセス
23
サーバプロセス
21
ストリーム処理部
TCPパケット処理層
132
IPパケット処理層
131
パケットフィルタ
111
113
転送制御表
112
TCPパケット処理層
1242
IPパケット処理層
1241
外部ネットワークドライバ
114
内部ネットワークドライバ
115
内部ネットワークドライバ
22
5
3
クライアントへ
ノード間結合ネットワーク

【図６】

転送制御表と割当済転送経路一覧表の説明図

（ａ）転送制御表の説明

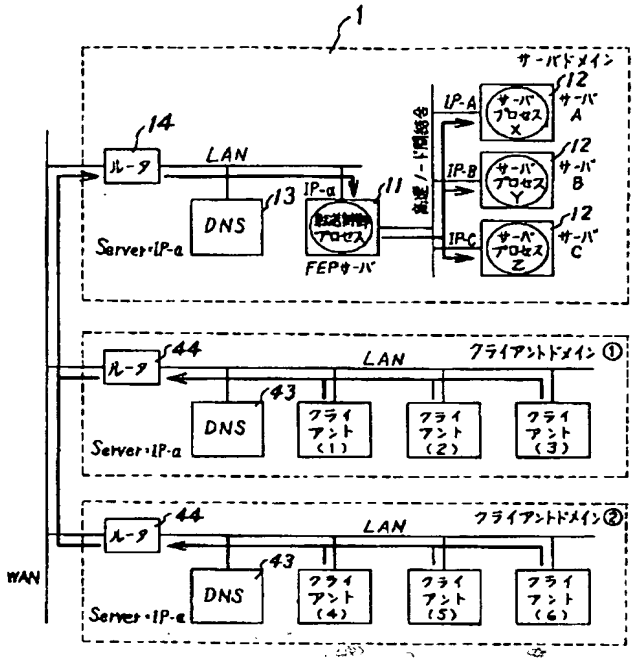
サービス識別子	転送先ＢＥＰノード識別子	比　率
デスティネーションポート番号	デスティネーションＩＰアドレス	
8080	ＩＰ－Ａ	1
	ＩＰ－Ｂ	1
8081	ＩＰ－Ｃ	2
	ＩＰ－Ｄ	1
8082	ＩＰ－Ｅ	1
	ＩＰ－Ｆ	1

（ｂ）割当済転送経路一覧表の説明

サービス識別子	クライアント識別子		転送先ＢＥＰノード識別子
デスティネーションポート番号	ソースＩＰアドレス	ソースポート番号	デスティネーションＩＰアドレス
8080	IP-CL1	PORT-CL1-ex1	IP-A
	IP-CL1	PORT-CL1-ex2	IP-B
	IP-CL2	PORT-CL2-ex1	IP-A
	IP-CL3	PORT-CL3-ex1	IP-B
8081	IP-CL4	PORT-CL4-ex1	IP-C
	IP-CL5	PORT-CL5-ex1	IP-D
	IP-CL6	PORT-CL6-ex1	IP-C
	IP-CL7	PORT-CL7-ex1	IP-C
	IP-CL8	PORT-CL8-ex1	IP-D
	IP-CL9	PORT-CL9-ex1	IP-C
8082	IP-CL10	PORT-CL10-ex1	IP-E
	IP-CL1	PORT-CL1-ex3	IP-F
	IP-CL2	PORT-CL2-ex2	IP-E
	IP-CL3	PORT-CL3-ex2	IP-F

【図７】

従来例の説明図（１）



【図８】

従来例の説明図（２）

（ａ）仕事の割り振り表の説明

サービス識別子	転送先サーバ識別子
サービスＸ	サーバＡ
サービスＹ	サーバＢ
サービスＺ	サーバＣ

（ｂ）ＩＰアドレス例の説明

サービス名	ＩＰアドレス
ftp-server	133.160.12.5